

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:47 ; Search time 91.75 Seconds

(without alignments)
15.135 Million cell updates/sec

Title: US-09-331-631a-5_COPY_33_75

Perfect score: 248
Sequence: 1 NOEDPQTECQOCQRRCKROOE.....RQOYQCRCKEICEEEREY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	42.3	588	1	VCIA_GOSHI
2	94	37.9	605	1	VCIA_GOSHI
3	71	28.6	47	1	AGRP_LOFCY
4	68	27.4	154	1	YS1_CAEEL
5	65.5	26.4	33	1	MBPI_MAIZE
6	65	26.2	919	1	ANDR_HUMAN
7	64	25.8	648	1	KAPC_DICDI
8	64	25.8	911	1	ANDR_PANTR
9	63.5	25.6	50	1	HSP1_MOUSE
10	63.5	25.6	50	1	HSP1_MOUSE
11	63	25.4	626	1	AH12_ARAHY
12	61	24.6	223	1	CAS2_SHEEP
13	61	24.6	429	1	AP44_MACFA
14	61	24.6	907	1	ANDR_CANFA
15	60	24.0	285	1	INVO_CANFA
16	59.5	24.0	614	1	AH11_ARAHY
17	59	23.8	758	1	YM38_YEAST
18	59	23.8	2124	1	Y192_HUMAN
19	58.5	23.6	639	1	GLCX_SOYBN
20	58.5	23.6	1407	1	TRHY_RABIT
21	58	23.4	223	1	CAS2_CAPHI
22	58	23.4	644	1	BTG_DROME
23	58	23.4	2318	1	NTC1_MOUSE
24	57.5	23.2	338	1	FSH_HUMAN
25	57.5	23.2	605	1	GLCA_SOYBN
26	57.5	23.2	991	1	DHPI_SCHPO
27	57	23.0	210	1	YOEK_ECOLI
28	57	23.0	1898	1	TRHY_HUMAN
29	57	23.0	1905	1	TAGB_DICDI
30	56.5	22.8	284	1	TPM1_RAT
31	56.5	22.8	284	1	TPM2_HUMAN
32	56.5	22.8	284	1	TPMB_HUMAN
33	56.5	22.8	284	1	TPMB_RABIT

34	56	22.6	239	1	CAUD_MELGA
35	56	22.6	524	1	SBP_SOYBN
36	56	22.6	771	1	CAUD_CHICK
37	55.5	22.4	49	1	HSP1_PIG
38	55.5	22.4	174	1	GRE_METTI
39	55.5	22.4	284	1	TPM1_CHICK
40	55.5	22.4	284	1	TPM1_MOUSE
41	55.5	22.4	284	1	TPMB_CHICK
42	55.5	22.4	303	1	BNK_DROME
43	55.5	22.4	441	1	PPAS_HUMAN
44	55	22.2	339	1	TF2D_HUMAN
45	55	22.2	450	1	SRF_DROME

ALIGNMENTS

RESULT	ID	VCIA_GOSHI	STANDARD	PRT	588 AA.
AC	P09801	VCIA_GOSHI	STANDARD	PRT	588 AA.
DT	01-MAR-1989	(Rel. 10, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots: Rosidae: eurosids II: Malvales: Malvaceae: Gossypium.				
RN	[1]				
RA	SEQUENCE FROM N.A.				
RP	Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVII. cDNA and amino acid sequences of the members of the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLICITIN, ETC.).				
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL: M16891; AAA33071.1; -				
DR	PIR: A30838; FMCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -				
DR	PFAM: PF00546; Seedstore_7s; 1.				
FT	Seed storage protein; Signal.				
FT	SIGNAL 1 25				
FT	CHAIN 26 588				
FT	SEQUENCE 588 AA: 69729 MW: 636599B29AB8ADEB CRC64;				
QY	3	EDPQTECQOCQRRCKROOE	SDPQOYQCRCKEICEEEREY	42	
DB	81	EDPQTECQOCQRRCKROOE	SDPQOYQCRCKEICEEEREY	118	
RESULT	2	VCIA_GOSHI	STANDARD	PRT	605 AA.
VCIA_GOSHI		STANDARD	PRT	605 AA.	

CC -1- MASS SPECTROMETRY: MM=5693.39; METHOD=MALDI.

CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.

RL Biosci. Biotechnol. Biochem. 61:984-988(1997).

RT "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from

RA Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;

RX MEDLINE=97357433; PubMed=9214759;

OC Eukaryota. Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

OC Cucurbitales; Cucurbitaceae; Luffia.

RN [1]

RP SEQUENCE.

RC TISSUE=SEED:

RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;

RT "Developmental biochemistry of cottonseed embryogenesis and

RT germination. XIX. Sequences and genomic organization of the alpha

RT globulin (vicillin) genes of cottonseed."

RL Plant Mol. Biol. 9:533-546(1987).

CC -1- FUNCTION: SEED STORAGE PROTEIN.

CC -1 SUBCELLULAR LOCATION: CYTOPLASM; MEMBRANE-BOUND VACUOLAR PROTEIN

CC BODIES.

CC -1 SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,

CC CONVICILIN, CONVELYCININ, ETC.).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or_send_email_to_license@isb-sib.ch)

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M19378; AAA33069.1; .

DR PIR: S06398; S06398.

DR HSSP: P50477; ICAX.

DR INTERPRO: IPR001113; .

DR PfAM: PF00546; Seedstore_7s; 1.

KW Seed storage protein; Signal.

FT SIGNAL

FT CHAIN 1 23

FT FT 24 605 VICILIN GC72-A.

SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

QY Query Match 37.9%; Score 94; DB 1; Length 605;

Best Local Similarity 35.9%; Pred. No. 0.0074;

Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 115 QDDPTGECOCORCRROESPDPGOGYOVCRRCKEICEEE 40
I : I : ::::II: II : :: I I : I :
115 QQQPKGFKECGQRQMGDEGRPERKRGCVKCEGRDYQGD 153

RESULT 3

AGRP_LUFCY STANDARD; PRT; 47 AA.

ID AGRP_LUFCY

P56568:

DC 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE 6.5 kDA ARGinine/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).

OS Luffia cylindrica (Smooth loofah) (Sponge gourd).

CS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

OC Cucurbitales; Cucurbitaceae; Luffia.

RN [1]

RP SEQUENCE.

RC TISSUE=SEED:

RA MEDLINE=97357433; PubMed=9214759;

RX Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;

RT "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from

RT the seeds of sponge gourd (luffia cylindrica)";

RL Biosci. Biotechnol. Biochem. 61:984-988(1997).

CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON

CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.

```
CC      -1- SIMILARITY: SOME, TO 7S SEED STORAGE PROTEINS.
KM Seed storage protein.
FT DISULFID 12 33
FT DISULFID 16 29
SQ SEQUENCE 47 AA: 5698 MW: 588B0FC82273AC05 CRC64;

Query Match          28.6%; Score 71; DB 1; Length 47;
Best Local Similarity 37.1%; Pred. NO. 0.18;
Matches 13; Conservative 9; Mismatches 9; Indels 4; Gaps 1;

OY      5 POTECCOCORRCROESDPDQOQCVRCKEICE 39
       ||| : | : | : | ||| : ||| : ||| :
DB      5 PRTEYACRVCQAHEGVER----QRRCQQVCEK 35

RESULT 4
YS5L CAEEL
ID YS5L CAEEL STANDARD: PRT: 154 AA.
AC Q09662;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHELTICAL 15.6 KDA PROTEIN ZK673.1 IN CHROMOSOME II PRECURSOR.
GN ZK673.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
[1]
RP STRAIN-BRISTOL NZ;
RA McMurtry A.;
RL Submitted (Mar-1995) to the EMBL/Genbank/DBJ databases.
[2]
REVISIONS.
RC STRAIN-BRISTOL NZ;
RA Jones S.J.M.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
-1- SIMILARITY: STRONG, TO C.ELEGANS E04D5.4 AND SOME, TO C.ELEGANS
    CC C03G6.13.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC DR EMBL; Z48585; CAA8486.1; -.
DR HSSP; P01551; IACX.
DR WORMPEP; ZK673.1; CE16742.
KW Hypothetical protein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 154 POTENTIAL.
SQ SEQUENCE 154 AA: 15637 MW: 323A0ZC3A2E782AB CRC64;

Query Match          27.4%; Score 68; DB 1; Length 154;
Best Local Similarity 35.5%; Pred. No. 0.96;
Matches 11; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

OY      3 EDPTCCOCORRCROESDPDQOQCVRCKEICE 33
       ||| : | : | : | ||| : ||| : ||| :
DB      69 DDPNTDCIQYTSLSNAKYTPILDQFCFKTC 99

RESULT 5
MBP1 MAIZE
ID MBP1 MAIZE STANDARD: PRT: 33 AA.
AC P28794;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
```



```

RC STRAIN=C129;
RA Schlueter G., Engel W.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBUNIT: CROSS-LINKED BY INTERCHAIN DISULFIDE BONDS AROUND THE
CC DNA-HELIX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K02926; AAA39980.1; -
DR EMBL: X07625; CAA30472.1; -
DR EMBL: X14003; CAA32169.1; -
DR EMBL: M27500; AAA39985.1; -
DR EMBL: Z47352; CAA87410.1; -
DR PIR: A02660; HSMMS1.
DR PIR: S03820; S03820.
DR PIR: A28331; A28331.
DR MGD: MGI:97765; PRL1.
DR INTERPRO: IPR000221; -.
DR PFAM: PF00260; protamine.P1.1.
DR PROSITE: PS00048; PROTAMINE.P1.1.
DR Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
FT DISULFID 5 INTERCHAIN (WITH CYS-21) (BY SIMILARITY).
FT DISULFID 6 BY SIMILARITY.
FT DISULFID 21 INTERCHAIN (WITH CYS-5) (BY SIMILARITY).
FT DISULFID 36 INTERCHAIN (WITH CYS-36) (BY SIMILARITY).
FT DISULFID 37 BY SIMILARITY.
SQ SEQUENCE 50 AA; 6827 MW; B61CAF3D1BBF4978 CRC64;

Query Match 25.6%; Score 63.5; DB 1; Length 50;
Best Local Similarity 28.9%; Pred. No. 1.1;
Matches 11; Conservative 11; Mismatches 11; Indels 5; Gaps 1;

Oy 6 QTECCQOCRRCRQESDPROQCYCQRCKEICEEEY 43
Db 11 RSRRCRRRRRCRR-----RRRCRRRRRRRCRRRSY 43

RESULT 10
HSP1_RAT
ID HSP1_RAT STANDARD: PRT; 50 AA.
AC P10118:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPERM PROTAMINE P1 (CYSTEINE-RICH PROTAMINE).
GN PRM1 OR PRM-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89335257; PubMed=2757789;
RA Klemm U., Lee C.H., Buitend P., Hake S., Engel W.;
RT "Nucleotide sequence of a cDNA encoding rat protamine and the haploid
RT expression of the gene during rat spermatogenesis.";
RL Biol. Chem. Hoppe-Seyler 370:293-301(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96341725; PubMed=8720108;

```

```

RA Schlueter G., Celik A., Obata R., Schlicker M., Hofferbert S.,
RA Schlung A., Adham I.M., Engel W.;
RT "Sequence analysis of the conserved protamine gene cluster shows that
RT it contains a fourth expressed gene.";
RL Mol. Reprod. Dev. 43:1-6(1996).
RN [3]
RP SEQUENCE.
RX MEDLINE=89207111; PubMed=3072011;
RA Ammer H., Henschen A.;
RT "Rat sperm protamine. Isolation and sequence analysis.";
RL Biol. Chem. Hoppe-Seyler 369:1301-1306(1988).
RN [4]
RP SEQUENCE OF 1-15 AND 44-50.
RX MEDLINE=76184789; PubMed=1268226;
RA Kistler W.S., Keim P.S., Heinrichson R.L.;
RT "Partial structural analysis of the basic chromosomal protein of rat
RT spermatozoa.";
RL Biochim. Biophys. Acta 427:752-757(1976).
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBUNIT: CROSS-LINKED BY INTERCHAIN DISULFIDE BONDS AROUND THE
CC DNA-HELIX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z46939; CAA87061.1; -
DR PIR: A27128; A27128.
DR PIR: S03997; S03997.
DR INTERPRO: IPR000221; -.
DR PFAM: PF00260; protamine.P1.1.
DR PROSITE: PS00048; PROTAMINE.P1.1.
DR Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MET 0
FT DISULFID 5 INTERCHAIN (WITH CYS-21) (BY SIMILARITY).
FT DISULFID 6 BY SIMILARITY.
FT DISULFID 21 INTERCHAIN (WITH CYS-5) (BY SIMILARITY).
FT DISULFID 36 INTERCHAIN (WITH CYS-36) (BY SIMILARITY).
FT DISULFID 37 BY SIMILARITY.
SQ SEQUENCE 50 AA; 6889 MW; B60B9F3D1BB6D978 CRC64;

Query Match 25.6%; Score 63.5; DB 1; Length 50;
Best Local Similarity 28.9%; Pred. No. 1.1;
Matches 11; Conservative 11; Mismatches 11; Indels 5; Gaps 1;

Oy 6 QTECCQOCRRCRQESDPROQCYCQRCKEICEEEY 43
Db 11 RSRRCRRRRRCRR-----RRRCRRRRRRRCRRRSY 43

RESULT 11
AH12_ARAHY
ID AH12_ARAHY STANDARD: PRT; 626 AA.
AC P43238:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ALLEGREN ARA H 1, CLONE P41B (ARA H 1).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Arachis.
RN [1]

```


KW	Plasma Lipid transport; HDL; Chylomicron; Repeat; Signal.
FT SIGNAL	1 20 BY SIMILARITY.
FT CHAIN	21 429 APOLOPROTEIN A-IV.
FT DOMAIN	33 330 13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT	33 34 1.
FT REPEAT	60 81 2.
FT REPEAT	82 103 3.
FT REPEAT	115 136 4.
FT REPEAT	137 158 5.
FT REPEAT	159 180 6.
FT REPEAT	181 202 7.
FT REPEAT	203 224 8.
FT REPEAT	225 246 9.
FT REPEAT	247 268 10.
FT REPEAT	269 286 11.
FT REPEAT	287 308 12.
FT REPEAT	309 330 13.
FT DOMAIN	372 420 GLU/GLN-RICH.
SQ SEQUENCE	429 AA; 49876 MW; 3d45bf551d0db60c CRC64;

```

Query Match          24.6% Score 61: DB 1; Length 42;
Best Local Similarity 34.1%; Pred. No. 11;
Matches 14; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

QY      2 QEDPQTECCGCCRCRCRCQESDPDQDQCYCRCKEICEEEEE 42
      1 - - - - - : : : : : : : : : : : : : : : : :
Db      376 QREQQQEEQQQEEQQEEQQQEEQQQEEQQQEEQQQEEQQ 416

RESULT 14
ANDR_CANFA
ID  ANDR_CANFA      STANDARD:      PRT:      907 AA.
AC  09T790:
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DE  30-MAY-2000 (Rel. 39, Last annotation update)
DE  ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
GN  AR OR NR3C4.
OS  Canis familiaris (dog).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Lu B., Smock S.L., Castleberry T.A., Owen T.A.:
RT  "Molecular cloning and functional characterization of the canine
RT  androgen receptor."
RL  Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC  THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC  PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC  -1- SUBCELLULAR LOCATION: NUCLEAR.
CC  -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC  A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC  -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC  NR3 SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AF197950; AAF18084.1; -
DR  PROSITE: PS009031; NUCLEAR_RECEPTOR; 1.
KW  Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW  Zinc-finger; Steroid-binding.
FT  DOMAIN 1 545 MODULATING (BY SIMILARITY).
FT  DNA_BIND 547 612 C4-TYPE ZINC FINGERS (TWO).
FT  ZN_FING 547 567 C4-TYPE.
FT  ZN_FING 583 607 C4-TYPE.

```

FT	DOMAIN	678	907	LIGAND-BINDING.
FT	DOMAIN	55	64	POLY-GLN.
FT	DOMAIN	70	76	POLY-GLN.
FT	DOMAIN	131	134	POLY-GLN.
FT	DOMAIN	180	202	POLY-GLN.
FT	DOMAIN	329	332	POLY-SER.
FT	DOMAIN	375	384	POLY-PRO.
FT	DOMAIN	399	405	POLY-ALA.
50	SEQUENCE	907 AA:	98726 MW:	C6819P780DD338AP CRC64

	Query Match	24.6%	Score 61:	DB 1:	Length 907:
	Best Local Similarity	48.1%	Pred.	No. 20:	
	Matches 13:	Conservative	5:	Mismatches	9;
				Indels	0; Gaps
Oy	5 PQTETCQCQCRRCRQRQESDPDPRQQYCYQR	31			
	: : :	: : : : : : : :			
Db	50 PGAILQLQKDKKKKKKKKKSTSPFRGKKKKKKKK	76			

INVO_CANFA	STANDARD:	PRT:	285 AA.
AC	P18174.		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-NOV-1990 (Rel. 16, Last sequence update)		
DT	01-FEB-1996 (Rel. 33, Last annotation update)		
DE	INVOLUCRIN.		
GN	IVL.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90348475; Pubmed=2385171;		
RA	Tseng H., Green H.;		
RT	"The involucrin genes of pig and dog: comparison of their segments of		
RL	repeats with those of prosimians and higher primates.";		
SL	Mol. Biol. Evol. 7:293-302(1990).		
CC	-I- PUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS		
CC	IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO		
CC	MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE		
CC	FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.		
CC	-I- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND		
CC	OTHER STRATIFIED SQUAMOUS EPITHELIA.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaborator		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-slb.ch/announce		
CC	or send an email to license@isb-slb.ch).		
CC	-----		
DR	EMBL; M34442; AAA30853.1; -.		
DR	INTERPRO; IPR002360; -.		
DR	PROSITE; PS00795; INVOLUCRIN; 1.		
KW	Keratinocyte; Repeat.		
QJ	SEQUENCE 285 AA; 33384 MW; DCE1BD889248BEA CRC64;		

	Query Match	24.2%	Score 60	DB 1	Length 285
	Best Local Similarity	31.7%	Pred. NO. 10		
	Matches 13	Conservative 9	Mismatches 19	Indels 0	Gaps 0
Qy	2	QEDPTECCQCCRCRCRQESDPFRQQYQCCRCRKEICEEEE	42		
Db	142	QOOEOEOSEKEKLEEDQKQEEELQOQEOQOQKKECECKHQE	182		

```
Search completed: March 1, 2001, 16:16:47
Job time: 427 sec
```


